Detection of Chemicals in Mixed, Two-Dimensional Raman Spectra

David Gillis*, Jacob Grun, and Jeffrey Bowles
Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375
Robert Lunsford, Sergei Nikitin, and Charles Manka
Research Support Instruments, Lanham, MA 20706

ABSTRACT

Recently, researchers at the Naval Research Laboratory have developed the SWORrD system for measuring two-dimensional Raman Spectra. The device consists of a tunable 2d ultraviolet laser that illuminates the sample at various wavelengths (210-300 nm) and collects a single Raman spectrum at each laser wavelength. The single spectra are combined to form a two-dimensional spectrum (laser wavelength by scattered wavenumber).

In this paper we introduce a novel method for the detection of known agents ('targets') within measured 2d spectra. Our method is bases on 'linear mixed pixel' techniques from hyperspectral imagery; in particular, we generalize the Adaptive Subspace Detector (ASD) to a form suitable for SWORrD samples. Our detector uses the individual laser runs to define a set of points within wavenumber space; the set of points corresponding to a 2d spectra defines a particular subspace that contains each material. These subspaces are then used with ASD to identify targets. We include experimental results using real-world data to illustrate our results.

.

Keywords: Raman spectroscopy, Target Detection, Mixed Samples

1. INTRODUCTION

The Swept-Wavelength Optical resonance-Raman Device (SWOrRD) [1,2] is a new system that has been developed by the Naval Research Laboratory to measure resonant Raman spectrum over a variety of input laser wavelengths, ranging from the deep UltraViolet (UV) to the visible. SWOrRD consists of a tunable laser that illuminates the sample under study and a two-stage spectrometer for recording the Raman-scattered light that is emitted. The laser is a gain-switched Ti:Sapphire oscillator that operates at 5 kHz and generates 18 ns TEM-00 pulses tunable from 700 – 940 nm in 1 nm steps. Light from the oscillator is converted with barium borate crystals to either the third or fourth harmonic for output from 210 – 280 nm, with a spectral width of approximately 4 cm^{-1}. Tuning the laser is synchronized with the angular positions of the gratings in the spectrometer, and takes less than one minute to tune from one wavelength to the next.

By varying the input laser wavelength, SWOrRD creates a two-dimensional Raman spectrum by 'building up' a series of one-dimensional spectra, one for each laser input (Fig.1). Any input energy-dependent resonances will appear as 'bumps' within the SWOrRD spectrum. This combination of traditional Raman spectroscopy with resonance information creates a novel measurement method that contains much more information about a given sample than previous methods. Our goal is to use this extra information in order to develop new methods for identifying desired materials within unknown measured samples.

In this paper we introduce one method to for doing so, based on the 'linear mixing' assumption common in signal- and hyperspectral image processing. This model assumes that the spectrum corresponding to a given sample that includes several 'pure' components may be written (at least approximately) as a linear sum of the spectra of the individual components. Over the last several years, a wide variety of different 'target detection' algorithms [3] have been introduced in the literature to identify individual materials in a linear mixture. Our method is based on one such algorithm, known as the Adaptive Subspace Adapter (ASD). In very general terms, the ASD algorithm assumes that given target and background spectra can be modeled as subspaces in a certain *n*-dimensional vector space; and the decomposes an unknown sample into target / background components via linear algebraic projection operators. In short,

* Please address any correspondence to: David.Gillis@nrl.navy.mil

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
18 MAR 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Detection Of Chemicals In Mixed, Two-Dimensional Raman Spectra	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 Overlook Ave SW,Washington,DC,20706**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

NRL has developed the SWOrRD system for measuring two-dimensional Raman Spectra. The device consists of tunable 2d ultraviolet laser that illuminates the sample at various wavelengths and collects a single Raman spectrum at each laser wavelength. The single spectra are combined to form a two-dimensional spectrum (laser wavelength by scattered wavenumber). In this paper we introduce a novel method for the detection of known agents within measured 2d spectra. Our method is based on 'linear mixed pixel' techniques from hyperspectral imagery; in particular, we generalize the Adaptive Subspace Detector (ASD) to a from suitable for the SWOrRD system. Our detector uses the individual laser tune to define a set of points within wavenumber space; the set of points corresponding to a 2d spectra defines a particular subspace that contains each material. These subspaces are then used with ASD to identify targets.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**7**	

if the given sample contains a 'large' target component, it is assumed that sample contains the target material as one of the pure in the components in the mixture; if not, then the material is not present.

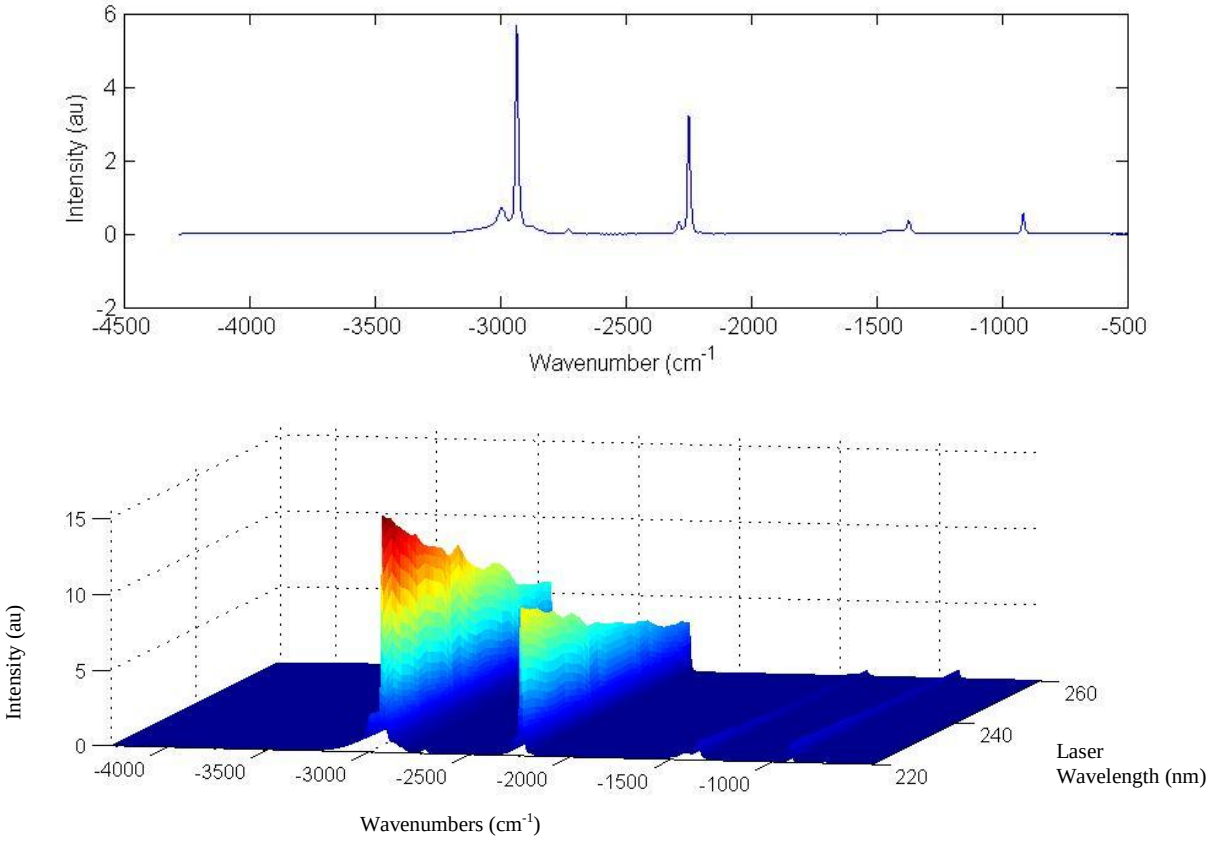


Fig. 1. SWOrRD Spectrum of Acetonitrile. Top: Raman spectrum from a single input illumination (240 nm). Bottom: The full 2d SWOrRD. The x axis corresponds to wavenumber; the y axis corresponds to the input laser wavelength.

In the next section, we present a general overview of the ASD algorithm, and discuss how we modified this algorithm to work with SWOrRD spectra. This is followed by the results of applying our modified algorithm to a variety of liquid and solid chemical mixtures.

2. TARGET DETECTION

In this section we present the details of our target detection algorithm. We begin this section with a brief overview of the linear mixing assumption, and how this assumption may be used to find targets in mixed spectra. This is followed by a general discussion of the Adaptive Subspace Detector (ASD), which forms the basis of our detection algorithm. We conclude by showing how we modified the ASD algorithm to work with the SWORRD data.

2.1 Linear Mixing

The main assumption underlying our detection algorithm is the Linear Mixing Model (LMM); intuitively, the linear mixing assumption is that a measured sample containing multiple materials can be written (at least approximately) as a weighted linear combination of the spectra of each of the components. In more formal terms, we begin by assuming that each sample spectrum **s** can be written as a *n*-dimensional vector $\mathbf{s} \in \Re^{\mathbf{n}}$. (For the moment, we will be deliberately

vague about what *n* represents; we will return to this question in Sec. 3.3.) If the sample contains the *k* 'pure' materials $\mathbf{s}_1, \ldots, \mathbf{s}_\mathbf{k}$, then the LMM assumption is that we can write **s** as the sum

$$\mathbf{s} = \sum_{\mathbf{i}=1}^{\mathbf{k}} \alpha_\mathbf{i} \mathbf{s}_\mathbf{i} + \eta \qquad (1)$$

where $\mathbf{s}_\mathbf{i} \in \Re^\mathbf{n}$ are the vectors corresponding to the individual components, $\eta \in \Re^\mathbf{n}$ is a noise / modeling error term, and $\alpha_\mathbf{i} \in \Re$ are scalars. Intuitively, the $\alpha_\mathbf{i}$ represent the 'fractional amount' of the material present in the sample, and are sometimes known as the *abundance* coefficients. Note that Eq. 1 can be written more succinctly in matrix form as

$$\mathbf{s} = \mathbf{M}\alpha + \eta$$

where **M** is the *n-by-k* matrix whose columns are the vectors $\mathbf{s}_\mathbf{i}$, and $\alpha = (\alpha_1, \ldots, \alpha_\mathbf{k})^T$ contains the abundances. With this notation, it can be shown that the least-squares optimal solution to estimating the abundance coefficients are given by the pseudo-inverse:

$$\hat{\alpha} = (\mathbf{M}^\mathbf{t} \cdot \mathbf{M})^{-1} \mathbf{M}^\mathbf{t} \cdot \mathbf{s} \qquad (2)$$

2.2 Mixed Spectra Target Detection

In the linear mixing model, it is assumed that a mixed sample spectrum can be decomposed into a sum of its constituents. In the target detection problem, the goal is to decide whether a given material (the 'target') is present in a given sample. To that end, we begin assuming that there exists a library of known 'background' spectra (that is, non-target materials that may or may not be present in a given sample), and, following Eq. 1, write a given sample **s** as the sum

$$\mathbf{s} = \sum_{\mathbf{i}=1}^{\mathbf{m}} \beta_\mathbf{i} \mathbf{s}_\mathbf{i} + \alpha \cdot \mathbf{t} + \eta$$

where $\mathbf{s}_\mathbf{i} \in \Re^\mathbf{n}$ are the background spectra, and $\mathbf{t} \in \Re^\mathbf{n}$ represents the target signature. In rough terms, the target detection problem reduces to deciding whether $\alpha = 0$ (which means the target is missing, or not present in the sample), or $\alpha \neq 0$ (target present).

Up to this point, we have been (implicitly) assuming that each component spectrum can be modeled as a single, unvarying vector $\mathbf{s}_\mathbf{i} \in \Re^\mathbf{n}$. In reality, different measurements of the same material will not be exactly identical, but will vary slightly from run to run, due to (among other things) differences in lab conditions, sample preparation, etc. In general, modeling this variability can be done either statistically (usually by assuming *a priori* some type of distribution) or geometrically. We use the second approach; in particular, we assume a *subspace* model for the variation. More precisely, we assume that, for each pure component, there exists a corresponding *n-by-l* matrix $\mathbf{S}_\mathbf{i}$ such that each measured component spectrum $\mathbf{s}_\mathbf{i}$ can be written as $\mathbf{s}_\mathbf{i} = \mathbf{S}_\mathbf{i}\sigma_\mathbf{i}$. Note that $\sigma_\mathbf{i}$ will vary among different measurements of the same material.

In the context of target detection, this means that there exists matrices **B** and **T** representing the background and target materials, respectively, such that every mixed sample **s** can be written as

$$\mathbf{s} = \mathbf{B}\beta + \mathbf{T}\alpha + \eta .$$

As above, the target detection problem reduces to deciding whether the components of α are zero (meaning the target is absent) or not (target present). More formally, this can be written as a statistical hypothesis test

$$\mathbf{H}_0 : \quad \mathbf{s} = \mathbf{B}\beta + \eta$$
$$\mathbf{H}_1 : \quad \mathbf{s} = \mathbf{B}\beta + \mathbf{T}\alpha + \eta$$

where the null hypothesis $\mathbf{H}_0$ states that the sample contains only background material (target absent), while the alternative hypothesis is that the target is present.

2.3 The Adaptive Subspace Detector

In target detection problems, the usual method for deciding hypothesis tests of the form Eq. 2 is the generalized likelihood ratio test (GLRT), in which the null hypothesis is rejected (that is, a target is deemed present) if and only if a certain statistic is above a given threshold. For the subspace model above, and assuming that the error is distributed as $\eta \sim \mathbf{N}(0, \sigma^2\mathbf{I})$, it can be shown that the GLRT statistic is given by

$$\mathbf{D}(\mathbf{s}) = \frac{\mathbf{s}^t(\mathbf{P}_\mathbf{B}^\perp - \mathbf{P}_\mathbf{Z}^\perp)\mathbf{s}}{\mathbf{s}^t\mathbf{P}_\mathbf{Z}^\perp\mathbf{s}}.$$

Here $\mathbf{s} \in \Re^\mathbf{n}$ is a measured sample, and $\mathbf{P}_*^\perp$ are (complementary) projection operators defined as

$$\mathbf{P}_\mathbf{B}^\perp = \mathbf{I} - \mathbf{P}_\mathbf{B} = \mathbf{I} - \mathbf{B}(\mathbf{B}^t\mathbf{B})^{-1}\mathbf{B}^t$$
$$\mathbf{P}_\mathbf{Z}^\perp = \mathbf{I} - \mathbf{P}_\mathbf{Z} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^t\mathbf{Z})^{-1}\mathbf{Z}^t$$

where $\mathbf{B}$ is the matrix of background signatures, and $\mathbf{Z} = [\mathbf{T} \mid \mathbf{B}]$ is the joint matrix of target and background signatures. The detector $\mathbf{D}$ is generally known as the adaptive subspace detector (ASD) in the signal- and hyperspectral image processing community, and as the F-test in statistics.

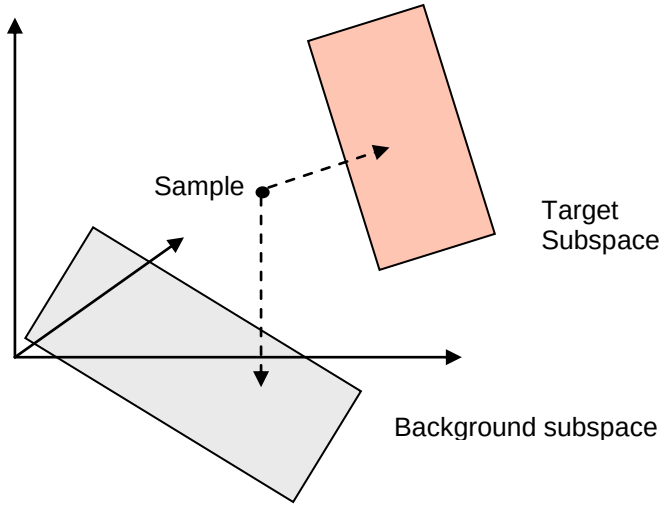


Fig. 2. Representation of the ASD algorithm. Each target spectrum and background spectra are used to define subspaces. Test cases are then projected into each subspace, and the distance from the subspace is calculated.

2.4 SWORRD ASD Target Detection

In order to implement the ASD target detector Eq. 3, it is necessary to define the target and background subspaces defined in the previous subsection. Implicit in this definition is how to represent the spectra as vectors in some *n*-dimensional space. Recall that a SWORRD spectrum is modeled as an *n-by-p* matrix, where *p* is the number of laser runs in the sample, and *n* is the number of bands (or wavenumbers) for each run. One way to model this would be as an *n x p* vector; however, this implies a tremendous number of dimensions (on the order of 10^6) for each vector. Such high-dimensionality has a number of unwanted consequences; for this reason, we use a different representation.

In geometrical terms, the ASD algorithm defines the space containing the data into target and background subspaces defined by the matrices **B** and **T**. Our goal is to define these subspaces; our approach is as follows: let **s** be a given 2d SWORRD spectrum, whose columns $\mathbf{s_i}$ are *n*-dimensional vectors corresponding to a single laser run (that is, a typical 1d Raman spectrum, taken at a certain laser wavelength). The SWORRD spectrum thus defines a set of points in *n*-dimensional wavenumber space; this set of points will define a relatively low-dimensional space (that can be determined via singular value decomposition, or SVD). We take this subspace to be the target / background subspaces needed in ASD.

To be more precise, suppose we are interested in detecting a certain material whose SWORRD spectrum is **t** against a background that is comprised of *m* SWORRD spectra $\mathbf{b}_1,\ldots,\mathbf{b}_\mathbf{m}$. We begin by calculating a 'target space' by modeling the individual runs $\mathbf{t}_1,\ldots,\mathbf{t}_\mathbf{p}$ as points in $\Re^\mathbf{n}$; we then use a standard SVD analysis to determine the dimensionality and basis vectors for the subspace containing these points. The basis vectors are then used to construct the *n-by-k* matrix **T** in Eq. 2 (here *k* is the estimated dimensionality of the target space). Similarly, we model each run $\mathbf{b_{i,j}}$ of background spectrum $\mathbf{b_i}$ a point in $\Re^\mathbf{n}$; the set of all *(m x p)* points $\mathbf{b}_{1,1},\mathbf{b}_{1,2},\ldots,\mathbf{b}_{1,\mathbf{p}},\mathbf{b}_{2,1},\ldots,\mathbf{b}_{\mathbf{m},\mathbf{p}}$ is again modeled via SVD to define a background matrix **B**. To run the detector against a sample **s**, we use the constructed matrices to apply Eq. 3 to each of the individual runs $\mathbf{s_i}$ of the sample **s**; each run outputs a scalar d_i that represents a 'target score' for that laser wavelength. To compute a final score D, we simply compute the norm of the individual outputs

$$\mathbf{D} = \sqrt{\mathbf{d}_1^{\,2} + \ldots + \mathbf{d}_\mathbf{p}^{\,2}} \ .$$

3. EXPERIMENTAL RESULTS

To test our algorithms, we began with a set of 5 liquid chemicals (acetonitrile, ethanol, methanol, ethylene glycol, and water), that were measured by SWORRD and used as our pure elements. We then created 14 various combinations of these chemicals, which were generally measured 3 or 4 times each, for a total of 68 samples. We also used a 'dummy' chemical (cyclohexane) that was included in our library but not in any of the mixtures; this was meant to test whether any 'false positives' would arise. An example of one of the mixture spectrum is shown in Fig. 3. The composition of the mixtures are summarized in table 1. We note that mixture consisted of equal volumes of each chemical present; using the molecular weight, we were then able to calculate the 'amount' (as measured by fractional molecular volume).

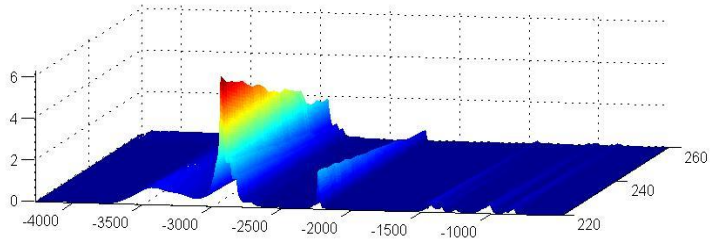


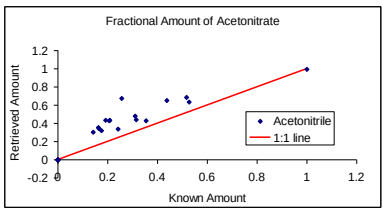
Fig. 3 SWORRD spectra from a liquid mixture of 4 chemicals (ethanol, methanol, ethylene glycol, and water)

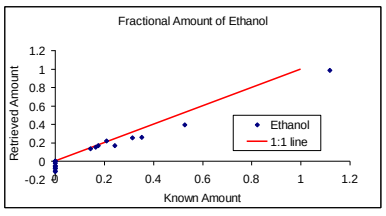
Our first experiment was to test whether the linear mixing model was appropriate for Raman spectral mixtures. In particular, we used the measured pure chemicals as the components in Eq. 1 above, and then 'unmixed', or calculated the abundances, for each mixtures using Eq. 2. The results are summarized in Fig. 4 below. Each chart shows the actual amount on the x-axis, and the estimated abundance on the y-axis; ideally, the samples would lie on the 1-1 line. As can be seen from the figures, the linear mixing model appears to be doing a pretty good job of estimating the actual concentrations, with the exception of water. We believe this is due to the fact that water has only very broad, not well-defined Raman features, and it is difficult to 'line up' the peak between different samples.

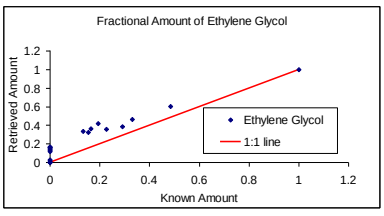
Our second experiment was to run the ASD detector described in Sec. 2. 4 above against each of samples. In the experiment, each individual chemical was defined as the 'target', and the remaining 5 chemicals (including cyclohexane) were assumed to be background; as a result, 6 different detectors were defined. The results are summarized in Fig. 5 below. Each figure shows the associated ASD score for the given chemical 'target' for each of the mixtures. For ease of interpretation, the scores are color-coded as follows: the green scores correspond to the pure samples, the blue scores are mixed samples which contain the associated target, and the red scores are samples which do not contain the target. It is easily seen that the ASD detector is able to find the target in each of the correct samples, and gives very low scares when the target is absent.

Sample	Components	Acetonitrile	Ethanol	Methanol	Ethylene Glycol	Water
1	2	✓			✓	
2	3	✓			✓	✓
3	3	✓		✓	✓	
4	4	✓		✓	✓	✓
5	2	✓	✓			
6	5	✓	✓	✓	✓	✓
7	4	✓	✓	✓		✓
8	3	✓	✓			✓
9	3	✓	✓		✓	
10	4	✓	✓		✓	✓
11	4	✓	✓	✓	✓	
12	2	✓		✓		
13	3	✓		✓		✓
14	2	✓				✓

Table 1. Composition of the 14 mixtures in the experiment. Each mixture contains equal volumes of each chemical.







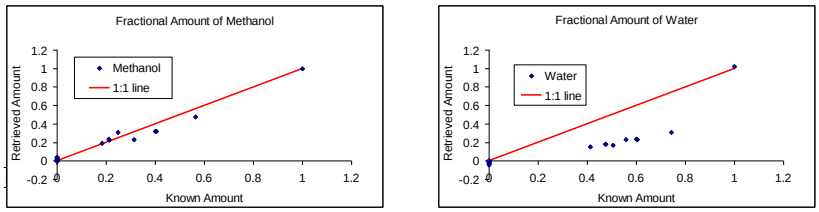


Fig 4. Actual vs. estimated amounts of each chemical in the mixtures. Clockwise from top left: acetonitrile, ethanol, ethylene glycol, water, and methanol.

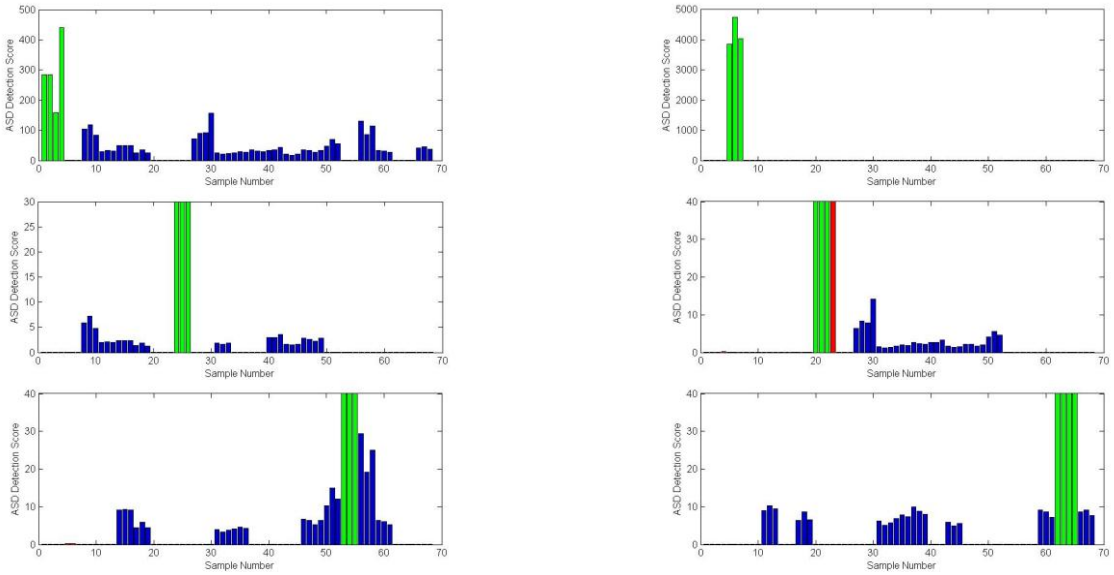


Fig. 5. ASD detector scores for each chemical and each sample. Top row: acetonitrile, cyclohexane. Middle row: ethylene glycol, ethanol. Bottom row: methanol, water.

ACKNOWLEDGEMENTS

This research was funded and supported by the US Defense Threat Reduction Agency (DTRA) and the US Office of Naval Research. Project number BA05DET019.

REFERENCES

1. J. Grun, et. al. "Identification of bacteria from two-dimensional resonant-Raman spectra", *Analytical Chemistry* 79 (14) 2007
2. G. Comanescu, et al. "Identification of explosives with two-dimensional ultraviolet resonance Raman spectroscopy", *Applied Spectroscopy* 62 (8) 2008
3. D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications", *IEEE Sig Proc Mag* 19 (1), 2002